SOBRE UNA CLASE DE MODELOS PARA LOS PROCESOS HIDROLOGICOS

Por Darío Valencia Restrepo
Profesor Asociado de la Universidad Nacional

INTRODUCCION

El creciente número de aplicaciones en el área de los recursos hidráulicos ha exigido un estudio más detenido de los fenómenos hidrológicos, dando origen durante la última década a una nueva disciplina conocida con el nombre de hidrología estocástica. Se ha tratado de contemplar las variaciones estocásticas que presentan dichos fenómenos, para que luego sea posible estimar la influencia que estas variaciones tienen sobre las consecuencias económicas y físicas de diversos proyectos hidráulicos.

La nueva metodología considera que los eventos hidrológicos ocurridos en el pasado, y los que tendrán lugar en el futuro, son funciones muestrales de procesos estocásticos muy complejos. Se deduce, entonces, que los futuros eventos hidrológicos estarán sujetos a las variaciones muestrales características de estos procesos, y de allí la importancia de modelos que proporcionen alguna idea sobre la naturaleza de estas variaciones.

Los modelos propuestos en el campo de la hidrología estocástica permiten generar series cronológicas, con frecuencia llamadas series sintéticas, que son estadísticamente indistinguibles de las series históricas en términos de ciertos estadísticos considerados relevantes. Ello es posible porque los parámetros de un determinado modelo son estimados, por regla general, utilizando directamente los estadísticos históricos. Pero como los estadísticos históricos son preservados por las series sintéticas solo como valores esperados (es decir, los estadísticos de estas últimas tienden a los históricos cuando la longitud de las series sintéticas es lo suficientemente grande), las series sintéticas finitas que se obtengan con el modelo estarán sometidas a variaciones muestrales, que se espera den una idea sobre las variaciones muestrales de las futuras series reales.

Claro que surge la pregunta sobre cuáles son los estadísticos "relevantes", dignos de ser preservados por las series sintéticas. Y es aquí en donde se dividen los especialistas de la nueva disciplina, surgen las diversas clases de modelos, debe señalarse los propósitos de cada modelo y el tipo de aplicación, y en donde existe un campo de trabajo muy abierto para los estudiosos.

Varias hipótesis gobiernan el diseño de los modelos para los principales procesos hidrológicos. La primera de ellas se refiere a la estacionaridad del proceso, es decir, a la suposición que las propiedades estadísticas del proceso no cambian en el tiempo (así por ejemplo, se dirá que el proceso que da origen a la serie de caudales anuales en un río es estacionario, si se supone que la media, la varianza, etc. de la serie permanecen invariables a lo largo del tiempo; suposición tal no puede aplicarse, como es fácil de ver, a la serie de caudales mensuales). Más tarde se verá cómo puede obviarse la no estacionaridad de ciertas series mediante la introducción de parámetros adicionales en el modelo correspondiente.

La segunda hipótesis, más fuerte que la primera y que en realidad la implica, tiene que ver con la llamada ergodicidad del proceso. Si se acepta esta hipótesis, los estadísticos de un determinado proceso pueden estimarse promediando a lo largo de una sola función muestral (la serie histórica, única disponible en el caso de los fenómenos hidrológicos), evitándose la necesidad de estimar efectuando el promedio sobre el conjunto de funciones muestrales (conjunto que en el caso de los fenómenos hidrológicos solo existe como abstracción de uso conceptual).

En este punto es importante destacar que los modelos aquí comentados son meramente descriptivos, en el sentido estadístico. No son modelos causales, ni modelos que pretendan explicar el carácter físico de los fenómenos hidrológicos.

Antes de cerrar la introducción, un ejemplo de mucha importancia actual permitirá ilustrar con una aplicación práctica el uso que puede hacerse de los modelos que han venido comentándose. Analizando el proyecto de desarrollo de una cuenca hidrográfica, se hace necesario someter a prueba el comportamiento (respuesta) de cada sistema (alternativa) de desarrollo que se proponga, con el fin de estimar su rendimiento ante las muy diferentes condiciones que puedan presentarse en el futuro. Si el sistema se somete solamente al estímulo de la serie de caudales históricos, se obtendrá apenas una respuesta individual del sistema. Pero como la serie histórica no va a repetirse nunca, es conveniente tener una estimación de la respuesta esperada ("promedia") del sistema y de la variabilidad de dicha respuesta, consecuencia de las posibles variaciones muestrales de las futuras series reales. Ello solo puede lograrse sometiendo el sistema al estímulo de diferentes series de entrada, algunas de ellas ojalá de duración muy larga. Y es aquí donde los modelos antes mencionados entran a cumplir una función, suministrando las series sintéticas correspondientes. En esta forma se procede para "simular" el comportamiento del sistema.

ALCANCES DE ESTE ARTICULO

Se presentará en este trabajo una clase muy general de modelos, introducida en [1], la cual incluye como casos particulares otros modelos antes propuestos en el campo de la hidrología estocástica. El modelo que representa a todos los miembros de la clase será estudiado con algún detenimiento, señalando sus propiedades esenciales y proporcionando condiciones suficientes para la existencia de *parámetros*. Especial atención se dedicará a un caso particular del *modelo general, también* introducido en [1], que da origen a los modelos llamados de desagregación, muy útiles en diferentes aplicaciones.

Inicialmente se describirán con brevedad algunos de los modelos antes propuestos, para luego presentar el modelo general, seguido del modelo de desagregación. Los interesados en los desarrollos matemáticos deberán consultar los apéndices que aparecen al final del artículo.

ALGUNOS MODELOS ANTERIORES

En 1954, Barnes [2] utilizó una tabla de números al azar para obtener una larga serie de caudales anuales que se asemejaba al registro histórico en términos de la media y la varianza, suponiendo una distribución normal de los caudales.

Thomas y Fiering, en 1962 [3], incluyeron además de la media y la varianza el coeficiente de correlación serial, por considerar que los registros históricos indicaban la importancia de preservar este último estadístico (consecuencia del fenómeno de persistencia observable en los procesos hidrológicos). El modelo propuesto por estos autores es de la forma

$$x_{t+1} = a\, x_t + b\, v_{t+1} \qquad (1)$$

en donde x_t : caudal en el período t

v_{t+1} : variable aleatoria independiente de x_t, con media cero y varianza unitaria, y que usualmente se supone distribuída normalmente.

a, b : parámetros que deben determinarse

Tomando adecuadamente valores esperados (esperanzas matemáticas) en la Ecuación (1), puede llegarse a que (ver Apéndice I)

$$\begin{aligned} a &= \rho \\ b &= \sigma \sqrt{1-\rho^2} \end{aligned} \qquad (2)$$

en donde ρ es el coeficiente de correlación serial y σ la desviación estándar de la serie de caudales (el modelo supone que la serie tiene media cero; no hay pérdida de generalidad porque ello siempre puede lograrse substrayendo la media de cada término de la serie original).

Utilizando los registros históricos se puede obtener estimados $\hat{\rho}$ y $\hat{\sigma}$ de los estadísticos en cuestión. Por tanto,

$$x_{t+1} = \hat{\rho}\, x_t + \hat{\sigma} \sqrt{1-\hat{\rho}^2}\; v_{t+1} \qquad (3)$$

es un modelo del proceso de caudales que preservará la media, la varianza y el coeficiente de correlación serial de los registros históricos (obsérvese que en el Apéndice I a y b se determinan de manera tal que el modelo preserve, como valores esperados, los estadísticos en cuestión).

Puede verse que el modelo dado por la Ecuación (3) es del tipo markoviano, pues el futuro (x_{t+1}) depende del presente (x_t), no interesando por cuál camino se llegó en el proceso a x_t. Varios especialistas sostienen que fenómenos hidrológicos como el de los caudales anuales en un río no son adecuadamente representados por modelos de dicho tipo, y proponen que se recurra a modelos de memoria más compleja que la markoviana, como los presentados en [4] y [5].

Para generar una primera serie sintética con ayuda del modelo dado por la Ecuación (3), bastará asumir un valor de x_t(por ejemplo, el último valor de la serie histórica, o la media de la serie histórica, o un número al azar obtenido con ciertos requisitos); generar un número al azar proveniente de una población normal con media cero y varianza unitaria, el cual será el valor de v_{t+1}; y efectuar los cálculos indicados por la ecuación para obtener el valor de x_{t+1}. Luego se repite el procedimiento, generando un nuevo número al azar, v_{t+2}, y utilizando a x_{t+1} como base para obtener a x_{t+2} mediante una nueva aplicación de la Ecuación (3), etc.

La necesidad de preservar propiedades de correlación espacial entre sitios vecinos de aforo, dio origen al desarrollo de modelos que generan caudales en varios sitios simultáneamente. Por el hecho de considerar diferentes variables a la vez, estos modelos se denominan multivariantes. En 1967, Matalas [6] propuso el siguiente modelo, en realidad una extensión del anterior al caso de series múltiples:

$$X_{t+1} = A\ X_t + B\ V_{t+1} \tag{4}$$

en donde

$$X_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot \\ \cdot \\ \cdot \\ x_{mt} \end{bmatrix}$$

es un vector de variables aleatorias, correspondientes a los caudales anuales en cada uno de m sitios considerados, para el período t.

$$V_{t+1} = \begin{bmatrix} v_{1,\,t+1} \\ v_{2,\,t+1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ v_{m,\,t+1} \end{bmatrix}$$

es un vector de variables aleatorias independientes entre sí e independientes de X_t, cada una de ellas con media cero y varianza unitaria. Es usual suponer que las $v_{i,\,t+1}$ se distribuyen normalmente.

A, B: matrices cuadradas de parámetros, cuyo orden respectivo es m, y que deben ser determinadas.

Conviene señalar que en este modelo, así como en el anterior, el valor en el período $t+1$ se obtiene como la suma de dos componentes: uno determinístico que es función del valor en el período anterior t, y otro estocástico que viene dado por $B\,V_{t+1}$.

Por el procedimiento explicado en los Apéndices I y II es posible encontrar estimados A y B de estas matrices, teniendo en cuenta que la Ecuación (4) es un caso particular de la (5), que se verá más adelante.

Las series sintéticas múltiples generadas mediante la Ecuación (4) preservarán para cada serie individual los mismos estadísticos contemplados por la Ecuación (1), y además preservarán la correlación espacial entre pares de sitios, para un mismo período o para períodos consecutivos.

EL MODELO GENERAL

Es posible observar que los modelos anteriores son casos particulares del modelo

$$Y = A\,X + B\,V \qquad (5)$$

en donde Y es un vector en columna de n variables aleatorias; X un vector en columna de m variables aleatorias; V un vector en columna de n variables aleatorias, cada una con media cero y varianza unitaria, independientes entre sí e independientes de X; A una matriz de parámetros, de orden n x m; y B una matriz de parámetros, de orden n x n.

Con un modelo como el dado por la Ecuación (5) se trata de representar el comportamiento de Y dado X, o en otras palabras, de proporcionar valores aceptables de Y cuando X es conocido. Es natural que el modelo en cuestión no permita representar todas las propiedades de que goza el comportamiento real de Y condicionado al valor de X, pues por una parte existen limitaciones estructurales inherentes al modelo, y por la otra, la forma como se determine las matrices A y B señalará qué propiedades preservará el modelo.

Como se explica en el Apéndice I, las matrices A y B pueden encontrarse mediante las ecuaciones

$$A = S_{YX}\,S_{XX}^{-1} \qquad (6)$$

$$B\,B^T = S_{YY} - S_{YX}\,S_{XX}^{-1}\,S_{XY} \qquad (7)$$

en donde S_{XX}, S_{XY}, S_{YX} y S_{YY} son las matrices de covarianza que pueden definirse para los vectores X y Y. Si se observa con cuidado los desarrollos del Apéndice I que conducen a las Ecuaciones (6) y (7), se verá que los parámetros A y B deben satisfacer esas ecuaciones si se desea que el modelo preserve todos los estadísticos incluídos en las matrices de covarianza S_{YX} $(= S_{XY}^T)$ y S_{YY}.

En la práctica, primero se estiman las matrices de covarianza utilizando los registros históricos disponibles de los vectores X y Y; luego, mediante la Ecuación (6) se obtiene un estimado de la matriz A; y resolviendo la Ecuación (7) para B (por un procedimiento como el indicado en el Apéndice II u otro cualquiera) se obtiene un estimado para B. Las series que el modelo suministra para Y preservarán todas las propiedades estadísticas de primero y segundo orden (es decir, medias, varianzas y coeficientes de correlación) asociadas con Y, y las relativas de Y a X.

Desde el punto de vista del análisis de regresión, se puede considerar a la Ecuación (5) como un conjunto de n ecuaciones del tipo

$$y_i = \sum_{j=1}^{m} a_{ij}\, x_j + \sum_{j=1}^{n} b_{ij}\, v_j \qquad (8)$$

$$i = 1, 2, \ldots, n$$

cada una de las cuales es el resultado de practicar una regresión múltiple de y_i sobre x_1, x_2, . . ., x_m; pero con la especial peculiaridad que la estructura de correlación interna de las y_i, dada por la matriz de covarianza S_{YY}, es mantenida a través de los elementos b_{ij} de la matriz B. En otras palabras puede decirse que las llamadas "perturbaciones", dadas por los segundos términos de las ecuaciones representadas por (8), deberán estar correlacionadas para que así pueda respetarse la correlación de las variables y_i.

MODELOS DE DESAGREGACION

$$X = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot \\ \cdot \\ \cdot \\ x_{mt} \end{bmatrix} = X_t$$

Para ilustrar la subclase de los modelos de desagregación introducidos en [1], puede considerarse la siguiente aplicación práctica. Sea X un vector cuyos elementos son las precipitaciones anuales en m sitios de una misma cuenca, para el año t.

$$Y = \begin{bmatrix} y_{11t} \\ y_{12t} \\ y_{13t} \\ y_{14t} \\ y_{21t} \\ y_{22t} \\ y_{23t} \\ y_{24t} \\ \cdot \\ \cdot \\ \cdot \\ y_{m4t} \end{bmatrix} = Y_t$$

Sea Y un vector cuyos elementos son las precipitaciones estacionales en los mismos sitios, durante el año t, suponiendo cuatro estaciones (dos secas y dos húmedas, por ejemplo). En este vector en columna, y_{ijt} denota la precipitación en el sitio i, durante la estación j del año t.

Por lo anterior, la matriz A resulta ser de orden 4m x m, la matriz B de orden 4m x 4m, y el vector en columna V de 4m elementos. Entonces, el modelo

$$Y_t = A\ X_t + B\ V_t \qquad (9)$$

permite obtener valores estacionales asociados con los valores anuales X_t. Estos últimos pueden conseguirse mediante un procedimiento autorregresivo como el dado por la Ecuación (4), o bien empleando modelos como los introducidos en [4] y [5]. Como se ve, el modelo de desagregación puede ser utilizado en conjunción con cualquier modelo que genere valores anuales. Los valores estacionales así obtenidos preservarán las medias y varianzas estacionales para cada sitio, así como la correlación existente entre valores estacionales de un mismo sitio o sitios diferentes (para un mismo año), y aquella existente entre cualquier valor estacional y cualquier valor anual (también dentro de un mismo año).

Además de lo anterior, los modelos de desagregación gozan de una propiedad muy importante, demostrada en el Apéndice III. En virtud de dicha propiedad, las cuatro precipitaciones estacionales generadas para un cierto sitio sumarán el valor anual correspondiente, respetando así una propiedad observable en los registros históricos. También puede decirse que dicha propiedad garantiza que la desagregación estacional respetará las propiedades preservadas por el modelo, cualquiera que fuere, empleado para generar los valores anuales (ésto resulta ser de especial importancia cuando para la generación anual se utiliza modelos destinados a preservar propiedades a largo plazo).

Mediante una nueva aplicación del modelo de desagregación, los valores estacionales ya generados pueden servir de base para conseguir valores mensuales correspondientes. En la Ecuación (5), Y sería ahora un vector de valores mensuales asociados con cierta estación, para los m sitios y en el año t; en tanto que X sería un vector de valores estacionales para los mismos sitios y en la misma estación del año t. Debe advertirse que ahora se requeriría un par de matrices A y B para cada estación del año.

CONCLUSIONES

Se ha puesto de presente la importancia de los modelos para los procesos hidrológicos, ya que ellos pueden proporcionar alguna idea sobre el carácter de las variaciones estocásticas de dichos procesos. Es posible que las series sintéticas obtenidas con los modelos suministren información sobre las variaciones muestrales que las series reales exhibirán en el futuro, de manera que se pueda estimar el comportamiento de sistemas sometidos a la incertidumbre de los fenómenos hidrológicos.

Todos los modelos estudiados contienen parámetros que usualmente se estiman por medio de los estadísticos de las series históricas. Los estadísticos de las series sintéticas tienden, como valores esperados, a los estadísticos de las series históricas, de manera que en este sentido los parámetros de los modelos son sesgados (recuérdese que los estadísticos históricos son apenas estimados de las propiedades probabilísticas de los procesos estocásticos considerados). El llamado enfoque bayesiano [9] permite mirar la cuestión en una forma diferente, pues define los estimados de los procesos como resultado de una combinación de la información muestral histórica con algún tipo de información a priori.

Ha sido tradicional emplear los modelos markovianos para las series estacionarias (anuales) e inclusive para series no estacionarias (estacionales, mensuales, etc.), pero en este último caso se presenta un serio problema al agregar los valores sintéticos obtenidos. En efecto, por ejemplo, la generación markoviana mensual no garantiza la preservación de importantes estadísticos a nivel anual. Para obviar esta cuestión, se introdujo un modelo de desagregación que permite, partiendo de valores anuales, obtener sucesivamente valores correspondientes a niveles más bajos de agregación. Se demuestra (ver Apéndice III) que el esquema de desagregación preserva los estadísticos en todos los niveles de agregación; en particular, la preservación de los estadísticos anuales hace muy atrayente el uso conjunto del modelo de desagregación con modelos no markovianos utilizados para el mantenimiento de efectos especiales (a largo plazo) en las series anuales.

También se señaló que el modelo de desagregación, lo mismo que los modelos markovianos, es un caso particular del modelo general, $Y = AX + BV$. Ello muestra el alto grado de flexibilidad proporcionado por el modelo general. Además, al estudiar detenidamente las propiedades principales del modelo general (preservación de momentos, relaciones con el análisis de regresión, estimación de parámetros, preservación de transformaciones lineales) quedan muy en claro las características principales de todos los modelos de la clase representada por la ecuación $Y = AX + BV$.

Finalmente, debe escribirse un breve comentario sobre el grado de fidelidad con que esta clase de modelos representa la realidad de los procesos considerados. La teoría de las distribuciones multivariantes indica que si los vectores aleatorios X y Y se distribuyen multinormalmente, entonces los modelos constituyen una representación exacta, en términos probabilísticos, de los procesos (ello explica por qué con frecuencia se buscan transformaciones de los datos históricos que acerquen su comportamiento al de números provenientes de una población multinormal). Sin embargo, las propiedades de primeros y segundos momentos preservadas por los modelos (medias, varianzas y covarianzas) se mantienen de todas maneras, en virtud de que la técnica de los valores esperados conduce a resultados independientes de las distribuciones subyacentes de X y Y.

APENDICE I

MATRICES DE VARIABLES ALEATORIAS: Definiciones y Propiedades

Considérese una matriz cuyos elementos son variables aleatorias

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1q} \\ x_{21} & x_{22} & \dots & x_{2q} \\ \dots & \dots & \dots & \dots \\ x_{p1} & x_{p2} & \dots & x_{pq} \end{bmatrix}$$

Por definición, el valor esperado de X es

$$E\,[X] = \begin{bmatrix} E\,[x_{11}] & E\,[x_{12}] & \dots & E\,[x_{1q}] \\ E\,[x_{21}] & E\,[x_{22}] & \dots & E\,[x_{2q}] \\ \dots & \dots & \dots & \dots \\ E\,[x_{p1}] & E\,[x_{p2}] & \dots & E\,[x_{pq}] \end{bmatrix}$$

Puede verse que E es un operador lineal, esto es, si X y Y son matrices de variables aleatorias, A, B y C matrices de constantes, y los productos $A \cdot X$ y $Y \cdot B$ están bien definidos, lo mismo que la suma $A X + Y B + C$, se tiene

$$E\,[AX + YB + C] = A\,E\,[X] + E\,[Y]\,B + C$$

Ahora supóngase que X y Y son vectores en columna

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

Denotando por D_X y D_Y los vectores de las medias, E [X] y E [Y] respectivamente, se concluye que

$$S_{XY} = E\,[\,(X - D_X)\,(Y - D_Y)^T\,]$$

es simplemente la matriz de covarianza entre los vectores X y Y. El elemento general (i, j) de esta matriz es la covarianza entre las variables x_i e y_j

$$S_{XY}\ (i, j) = E\ [\ (x_i - \mu_{x_i})(y_j - \mu_{y_j})\ ]$$

$$i = 1, 2, \ldots, m$$

$$j = 1, 2, \ldots, n$$

Si Y = X, S_{XX} es la matriz de covarianza de X consigo mismo. Los elementos de la diagonal principal de esta matriz son las varianzas de las variables x_i.

Suponiendo que las variables x_i son independientes entre sí, tienen media cero y varianza unitaria, puede escribirse que (independientemente de los tipos de distribuciones atribuíbles a las x_i)

$$E\ [\ X\ ] = O$$

$$E\ [x_i\, x_j] = \begin{cases} O & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

Entonces, $$S_{XX} = I$$

siendo I la matriz identidad de orden m.

ESTIMACION DE LOS PARAMETROS A Y B

En el modelo dado por la Ecuación (5)

$$Y = A\,X + B\,V$$

se presenta la necesidad de estimar los parámetros matriciales A y B. Para ello se empleará la técnica de los valores esperados o esperanzas matemáticas, que entre otras ventajas tiene la de ser independiente de los tipos de distribución asociados con las variables aleatorias consideradas.

Tomando directamente valores esperados en la ecuación,

$$\begin{aligned} E\ [\ Y\ ] &= E\ [\ A\,X + B\,V\ ] \\ &= A\ E\ [\ X\ ] + B\,E\ [\ V\ ] \end{aligned}$$

De manera que si las variables originales se normalizan previamente (por substracción de las respectivas medias) y si se supone que E [V] = 0, puede verse que

$$0 = 0 + 0$$

lo cual indica que el modelo preservará las medias de Y.

Ambos lados de la ecuación original pueden ser posmultiplicados por X^T para obtener

$$Y X^T = A X X^T + B V X^T$$

Tomando ahora valores esperados,

$$E [Y X^T] = A \, E [X X^T] + B \, E [V X^T]$$

o bien $$S_{YX} = A \, S_{XX}$$

puesto que la independencia entre V y X garantiza que $E [V X^T] = 0$.

De aquí sigue que

$$A = S_{YX} \, S_{XX}^{-1}$$

siempre que S_{XX} sea no singular.

Para determinar la matriz B, puede posmultiplicarse la ecuación original por Y^T para obtener

$$\begin{aligned} Y Y^T &= (A X + B V) (A X + B V)^T \\ &= (A X + B V) (X^T A^T + V^T B^T) \\ &= A X X^T A^T + A X V^T B^T + B V X^T A^T + B V V^T B^T \end{aligned}$$

Tomando ahora valores esperados, y teniendo en cuenta que $E [X V^T]$ y $E [V X^T]$ son matrices nulas por la independencia supuesta entre X y V, que $E [V V^T] = I$, y que $A = S_{YX} \, S_{XX}^{-1}$, se llega a

$$S_{YY} = S_{YX} \, S_{XX}^{-1} \, S_{XY} + B \, B^T$$

ó $$B \, B^T = S_{YY} - S_{YX} \, S_{XX}^{-1} \cdot S_{XY}$$

Esta es una ecuación matricial para B, que será resuelta en el Apéndice II.

Como el modelo de Thomas y Fiering, dado por la Ecuación (1), es un caso particular del modelo general, los parámetros a y b de aquél podrán encontrarse mediante

$$\begin{aligned} a = A &= S_{YX} \, S_{XX}^{-1} = (\rho \sigma^2)(\sigma^2)^{-1} \\ &= \rho \end{aligned}$$

$$b\,b^T = b^2 = B\,B^T = S_{YY} - S_{YX}\,S_{XX}^{-1}\,S_{XY} = \sigma^2 - (\rho\sigma^2)(\sigma^2)^{-1}(\rho\sigma^2)^T$$

$$\therefore \quad b = \sigma\,\sqrt{1-\rho^2}$$

APENDICE II

LA TECNICA DE LOS COMPONENTES PRINCIPALES

Las aplicaciones prácticas del modelo general dado por la Ecuación (5) requieren la resolución de una ecuación matricial del tipo

$$B\,B^T = C$$

en donde C es una matriz conocida (ver Apéndice I). La teoría matricial establece que C tiene que ser una matriz simétrica, semidefinida positivamente, si se desea que los elementos de B sean reales.

Antes de explicar cómo se aplica la llamada técnica de los componentes principales, es conveniente enunciar algunas propiedades de las matrices semidefinidas positivamente (ver Capítulos 9, 10 y 11 de [7]), como sería el caso de C.

a. Todos los autovalores de C son no negàtivos. Si C tiene orden n y rango s, entonces s autovalores son positivos y los restantes n – s son nulos.

b. Debido a que C es una matriz real y simétrica, C tiene n autovectores linealmente independientes. Denotando por P_1, P_2, ..., P_n estos autovectores (dispuestos como vectores en columna) una vez que hayan sido normalizados, se tiene que la matriz

$$P = [\,P_1\,P_2 \ldots P_n\,]$$

es una matriz ortogonal, esto es,

$$P^T\,P = P\,P^T = I$$

o bien

$$P_i^T\,P_j = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

c. Si r_1, r_2, ..., r_n son los autovalores de C (usualmente dispuestos en orden de magnitud, aunque pueden no ser todos distintos), y P_1, P_2, ..., P_n los correspondientes autovectores, se sabe que

$$C\,P_i = P_i\,r_i$$

$$C\,P = P\,R$$

en donde R es una matriz diagonal cuyos elementos son los autovalores de C. También,

$$C = P\ R\ P^{-1} = P\ R\ P^{T}$$

Ahora, para encontrar una solución de la ecuación $B\ B^T = C$, basta definir

$$B = P\ R^{1/2}$$

en donde $R^{1/2}$ es una matriz diagonal cuyos elementos son las raíces cuadradas de los autovalores de C. En efecto,

$$B\ B^T = (P\ R^{1/2})\ (P\ R^{1/2})^T \quad P\ R^{1/2}\ R^{1/2}\ P^T$$
$$= P\ R\ P^T = C$$

Si B_1 es cualquier matriz ortogonal (o sea $B_1\ B_1^T = I$) de orden n, entonces $B\ B_1$ es también solución de la ecuación que se viene estudiando. En efecto,

$$(B\ B_1)(B\ B_1)^T = B\ B_1\ B_1^T\ B^T = B\ B^T = C$$

lo cual prueba que la ecuación en cuestión no tiene solución única.

Otro procedimiento para resolver la ecuación puede encontrarse en [8], en donde se supone que B es una matriz triangular inferior.

Al comienzo de este Apéndice se dijo que la matriz C debe ser semidefinida positivamente. Como en las aplicaciones prácticas se tiene que

$$\hat{B}\ \hat{B}^T = \hat{C} = \hat{S}_{YY} - \hat{S}_{YX}\ \hat{S}_{XX}^{-1}\ \hat{S}_{XY}$$

debe señalarse un procedimiento para estimar estas matrices de covarianza de una manera tal que se obtenga una matriz $\hat{C}$ semidefinida positivamente. En el Capítulo IV de [1] se dan condiciones suficientes para que aquello se cumpla, indicando los estimadores que deben emplearse.

APENDICE III

PRESERVACION DE TRANSFORMACIONES LINEALES

Se va a demostrar que si X es una transformación lineal de Y (o sea $X = Q\ Y$, en donde Q es una matriz de constantes), esta transformación se preservará cuando una vez dado X se generen valores de Y mediante el modelo

$$Y = A\ X + B\ V$$

Antes de presentar la prueba, conviene discutir el sentido práctico que entraña esta propiedad. Para ello, se vuelve sobre el modelo de desagregación dado por la Ecuación (9), y se pone de presente que

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot \\ \cdot \\ \cdot \\ x_{mt} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \ldots & 0 & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_{11t} \\ y_{12t} \\ y_{13t} \\ y_{14t} \\ \cdot \\ \cdot \\ \cdot \\ y_{m4t} \end{bmatrix}$$

o sea $\quad X_t = Q\,Y_t$

lo cual expresa que la suma de las 4 precipitaciones estacionales en un sitio, para un año t, da el valor anual en el sitio. Como aquí podría aplicarse la propiedad citada anteriormente, se concluiría entonces que los valores estacionales generados mediante el modelo dado por la Ecuación (9) conservarían aquellas relaciones lineales con los correspondientes valores anuales.

Para empezar la demostración, se encontrarán expresiones para las matrices de covarianza S_{XX}, S_{XY} y S_{YX} teniendo en cuenta que $X = Q\,Y$.

$$S_{YX} = E\,[\,Y\,X^T\,] = E\,[\,Y\,(Q\,Y)^T\,]$$

$$= E\,[\,Y\,Y^T\,Q^T\,] = S_{YY}\,Q^T$$

$$S_{XY} = S_{YX}^T = Q\,S_{YY}$$

$$S_{XX} = E\,[\,(Q\,Y)\,(Q\,Y)^T\,]$$

$$= Q\,S_{YY}\,Q^T$$

Las ecuaciones para A y B, derivadas en el Apéndice I, se convierten ahora en

$$A = S_{YX}\,S_{XX}^{-1} = S_{YY}\,Q^T\,(Q\,S_{YY}\,Q^T)^{-1}$$

o sea que $\quad Q\,A = Q\,S_{YY}\,Q^T\,(Q\,S_{YY}\,Q^T)^{-1} = I$

$$B\,B^T = S_{YY} - S_{YX}\,S_{XX}^{-1}\,S_{XY}$$

$$= S_{YY} - S_{YY}\,Q^T\,(Q\,S_{YY}\,Q^T)^{-1}\,Q\,S_{YY}$$

o sea que $\quad Q\,B\,B^T\,Q^T = Q\,S_{YY}\,Q^T - Q\,S_{YY}\,Q^T\,(Q\,S_{YY}\,Q^T)^{-1}\,Q\,S_{YY}\,Q^T$

$$(Q\,B)\,(Q\,B)^T = 0$$

lo que implica $\qquad Q\,B = 0$

Si ahora en la Ecuación (9) se premultiplica por Q en ambos lados, se llega a

$$Q\,Y_t = Q\,A\,X_t + Q\,B\,V_t$$

$$= I\,X_t + O\,V_t$$

utilizando las dos relaciones antes obtenidas. De modo que

$$Q\,Y_t = X_t \qquad \text{Q. E. D.}$$

REFERENCIAS

[1] Valencia, D. y J. C. Schaake: *A Disaggregation Model for Time Series Analysis and Synthesis, Report No. 149, Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, 1972.*

[2] Barnes, F. B.: *Requirements for a City Water Supply, J. Inst. Engrs., Australia, No. 26, pág. 198, 1954.*

[3] Thomas, H. A. y M. B. Fiering: *Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation, Design of Water-Resource Systems, Capítulo 12, A. Maass y otros, Harvard University Press. 1962.*

[4] Mandelbrot, B. B. y J. R. Wallis: *Noah, Joseph, and Operational Hydrology, Water Resources Research, Vol. 4, No. 5, págs. 909 - 918, 1968.*

[5] Mejía, J. M., I. Rodríguez-Iturbe y D. R. Dawdy: *The Broken-Line Process as a Potential Model for Hydrologic Simulation, Water Resources Research, Vol. 8, No. 4, 1972.*

[6] Matalas, N. C.: *Mathematical Assessment of Synthetic Hydrology, Water Resources Research, Vol. 3, No. 4, págs. 937 - 945, 1967.*

[7] Noble, B.: *Applied Linear Algebra, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1969.*

[8] Young, G. K. y W. C. Pisano: *Operational Hydrology using Residuals, J. Hydr. Div., ASCE, Vol. 94, HY4, págs. 909 - 923, 1968.*

[9] Rodríguez-Iturbe, I., R. Lenton, J. Valdés y D. Valencia: *Bayesian Hydrologic Model Building, Trabajo presentado en el Simposio Internacional sobre Incertidumbre en los Sistemas Hidrológicos y de Recursos Hidráulicos, Universidad de Arizona, 1972.*